CSE 163 Data Visualization

Hunter Schafer



Some Data

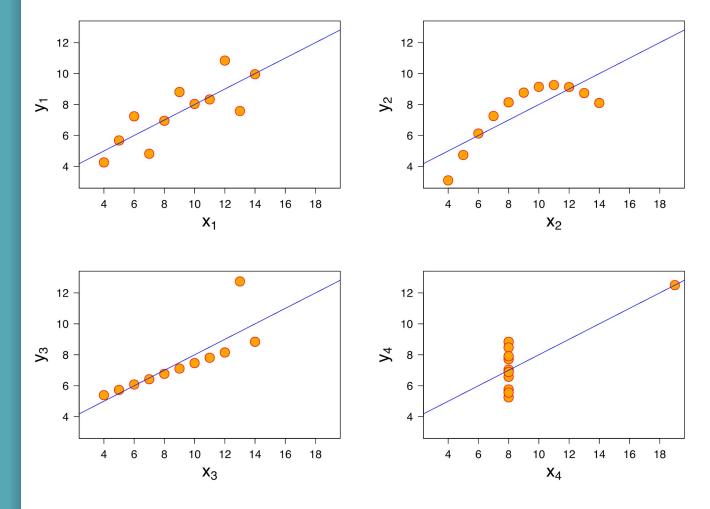
• Can you tell me the relationship between x and y?

$$\rho_{x} = 9.0 \quad \sigma_{x} = 3.317 \quad \mu_{y} = 7.5 \quad \sigma_{y} = 2.03$$
 $\rho_{y} = 3 + 0.5x \quad R^{2} = 0.67$

I		II		III		IV	
Х	у	Х	у	Х	у	Х	у
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

[Anscombe 73]

Some Data Visualizations

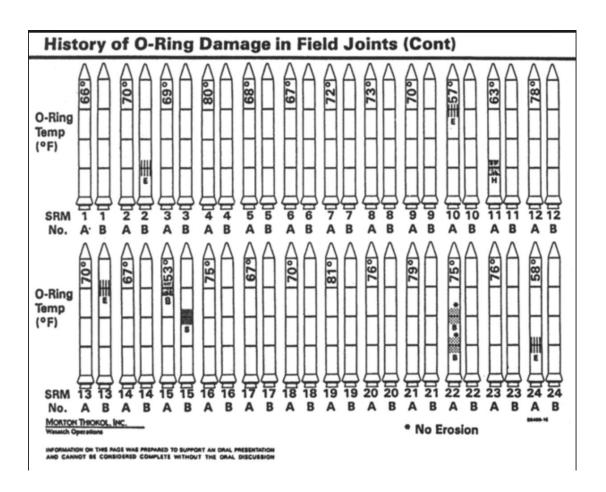


The Value of Visualization

- Answer questions (or discover them)
- Make decisioms
- See data im comtext
- Find patterns
- Present argument or tell a story
- Inspire

Make a Decision:

Challenger



Make a Decision:

Challenger



55°

Temperature (°F) of field joints at time of launch

O-ring damage

26°-29° range of forecasted temperatures (as of January 27, 1986) for the launch of space shuttle Challenger on January 28

35°



SRM 22

85°

In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]

Data in Context:

Cholera Outbreak



Data in Context:

Cholera Outbreak

Used map to hypothesize that pump on Broad St. was the cause. [from Tufte 83]



Data Viz in Python

- Many libraries exist to do graphics and plotting visualizations
- The most popular two are
 - matplotlib: extremely customizable (to a fault)
 - seaborn: great out-of-the-box visualizations)
- We will encourage using seaborn since it's quickly becoming the standard for data scientists for most of their work
 - We still will rely on matplotlib for some things
 - You are welcome to use it more in depth on your project to make highly customized charts

Reading Documentation

- Seaborn has excellent docs, great practice for reading!
- How to read documentation
 - 1. Skim examples, don't focus too much on code
 - 2. Read overview
 - 3. Look at examples and the code. Look at documentation for relevant parameters
 - 4. (Sometimes) Skim parameter list

- Here are the seaborn functions you might use
 - Bar/Violin Plot
 - Plot a Distribution
 - Scatter/Line Plot
 - <u>Linear Regression Plot</u>
 - Compare Two Variables
 - Heatmap

Seaborn Demo

- Working with a slightly different Pokemon dataset
- Visualize various aspects of the data
 - We only show a couple examples, but your homework asks you to do more!
- Remember to call sns.set()
- Here are the seaborn functions you might use
 - Bar/Violin Plot
 - Plot a Distribution
 - Scatter/Line Plot
 - Linear Regression Plot
 - Compare Two Variables
 - Heatmap



Visual Design Principles

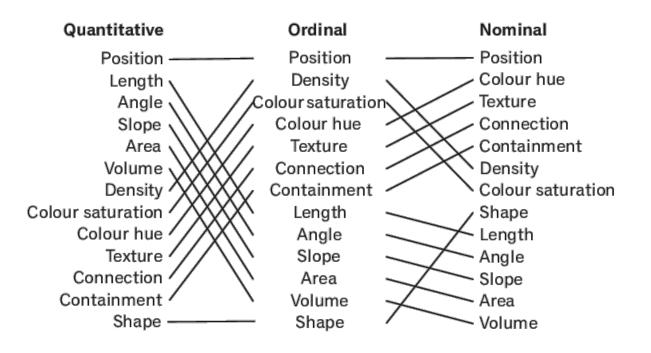
- Use effective encodings
- Avoid over-encoding
- Focus on readability, not aesthetics
- Avoid perceptual traps that can trick people in forming wrong conclusions

Rarely does a single visualization answer all questions. Instead, the ability to generate appropriate visualizations quickly is critical!

Types of Data

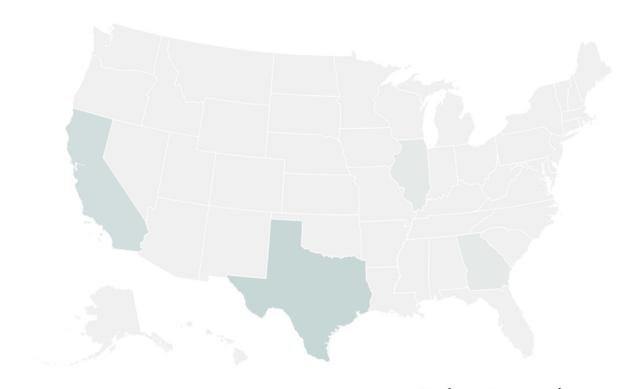
- Quantitative (numeric measurement)
 - Examples
 - Salary
 - age
- Ordinal (categorical, but there is an ordering)
 - Examples
 - First place, second place, third place
 - High School, Associates, Bachelors, ...
- Nominal (categorical, no ordering)
 - Examples
 - Type of Pokemon
 - School you attend

Encoding Effectiveness



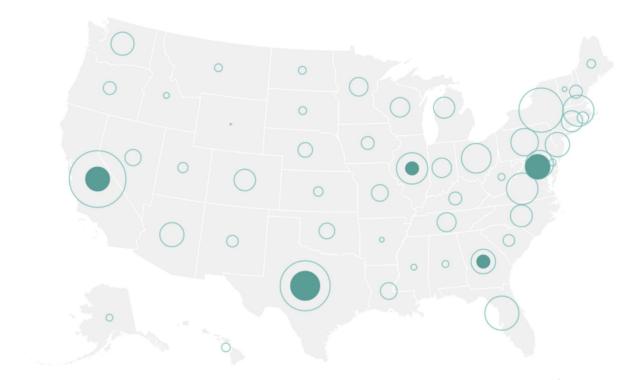
[Mackinlay 86]

Example Effectiveness



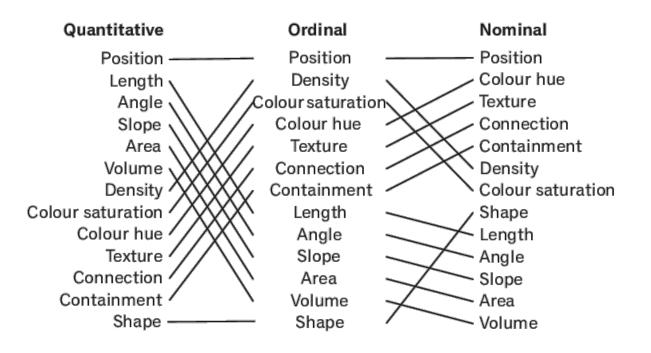
Color Encoding

Example Effectiveness



Area Encoding

Encoding Effectiveness

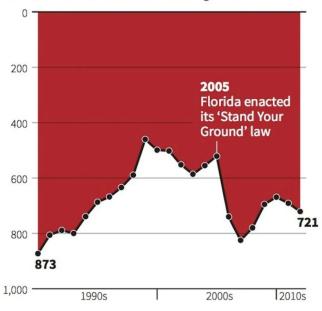


[Mackinley 86]

Readability

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014





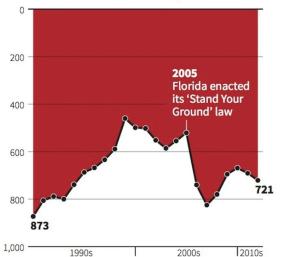


- Went up
- Went down
- About the same
- Don't remember

Yes, this was a real graphic posted in an article by Business Insider. They say it was made by Reuters.

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

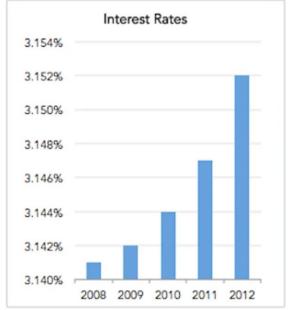
C. Chan 16/02/2014

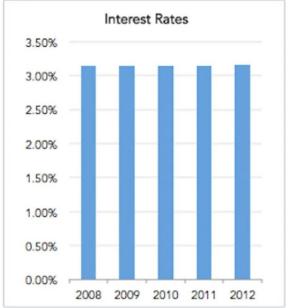




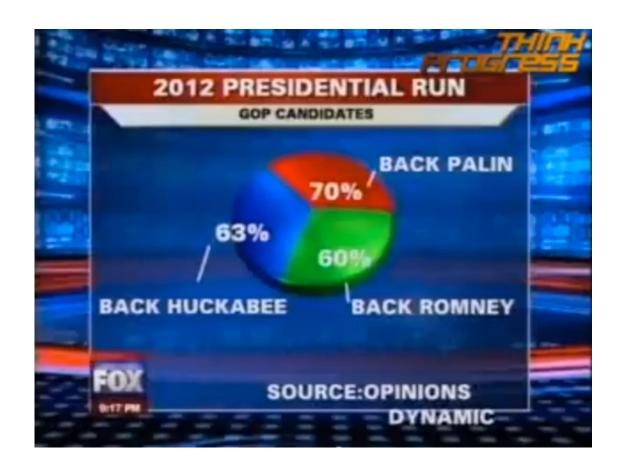
How to Lie with Data Viz

Same Data, Different Y-Axis





How to Lie with Data Viz



Seaborn

- Sometimes the plots come out not looking great, you will probably need to figure out how to fix them to make them clearer
- This can usually be solved by
 - Finding some more parameters in seaborn
 - Using matplotlib to save the day
- You also generally use matplotlib to get your plots to appear or to save them to a file



Default Parameters

 Python lets you define default values for parameters in case a client does not pass a specific value for that argument

```
def div(a=10, b=1):
    return a / b

div(2, 3) # 0.66
div(2) # 2
div(b=3) # 3.33
div() # 10
```

 This what allows libraries like seaborn or pandas to provide defaults!